

## Temas Selectos de Procesos Químicos-Análisis de Datos con Python

Clave: 2122075, Grupo: CH01

**Profesores:** Dr. José de Jesús Álvarez Ramírez ([jjar@xanum.uam.mx](mailto:jjar@xanum.uam.mx))  
M. en C. Abigail Marín López ([amarin@xanum.uam.mx](mailto:amarin@xanum.uam.mx))

**Tareas:** [temas.datosig@gmail.com](mailto:temas.datosig@gmail.com)

**Horario:** Lunes (AT-219), Miércoles (AT-106) y Viernes (C-105) de 12:00 a 13:30.

### 1. Objetivos del curso

Al finalizar, el alumnado será capaz de:

- Usar Python (Jupyter, NumPy, pandas) para **cargar, limpiar y manipular** datos experimentales.
- **Explorar y visualizar** datos de procesos (temperatura, presión, flujo, composición) y construir gráficos útiles para ingeniería (líneas, histogramas, diagramas de dispersión y mapas de calor).
- Analizar **series temporales de procesos** (remuestreo, ventanas móviles, tendencia y estacionalidad) aplicadas a datos de operación.
- Implementar **modelos básicos de ML** (regresión y clasificación) para problemas típicos de IQ: balance de energía, rendimiento, detección de condiciones anómalas.
- Desarrollar **un** proyecto aplicado de análisis de datos en un contexto de ingeniería química (secado, reactor, destilación, emisiones, etc.) priorizando reproducibilidad.

### 2. Programa y calendarización

#### Semana 1-2: Introducción a Python y entornos

- Jupyter, Python sintaxis básica, tipos de datos, estructuras de datos; lectura de CSV/Excel.
- Lectura de datos de laboratorio (destilación, secado, reacción).
- NumPy: matrices y computación vectorizada. (Álgebra de vectores y matrices)

#### Semanas 2-3: Manipulación y análisis de datos con pandas

- Indexación, filtrado, agrupación,
- Estadísticas descriptivas, unión de tablas.
- Pruebas de normalidad.
- Resumir, calcular la correlación y covarianza de las estadísticas descriptivas
- Visualización
- Acondicionamiento de datos
- Exploración de dataset sintético de balances de materia en una unidad (entradas/salidas por corriente);
- Exploración de datos de sensores T-P-F, porcentajes de conversión, selectividad, etc.

#### Semanas 4-5: Transformación de datos

- Datos de datos incompletos y duplicados,
- discretización, outliers, concatenación de datos, índices jerárquicos.

- Exploración de datos de reactor batch/pistón (T, t, composición).
- Outliers por fallas de sensor.
- Acondicionamiento y reporte automático (antes vs. después); bitácora de datos descartados y criterios de outlier (IQR/Z-score).

#### **Semanas 6–7: Visualización de datos**

- Gráficos con Matplotlib y Seaborn, diseño de figuras técnicas, mapas de calor
- Matplotlib/pandas/seaborn; diseño de figuras; barras/caja/lineas/dispersión/mapa de calor.
- Mapas de calor de operación de columna de destilación, secado y Rxn.

#### **Semanas 8–9: Series temporales**

- Fechas y frecuencias, remuestreo, ventanas móviles, tendencia/estacionalidad (STL), ciclos, detección simple de anomalías.
- Análisis de series temporales de procesos químicos.
- Reporte STL de señales de planta + tabla de indicadores (RMSE suavizado, amplitud estacional, % anomalías).

#### **Semanas 10–11: Introducción a ML y modelado**

- scikit-learn: split, escalado, regresión lineal, árboles de decisión, clasificación básica; validación y métricas (MAE, RMSE,  $R^2$ , accuracy, ROC AUC).
- **Modelación de series temporales de procesos químicos**
- baseline + un modelo mejorado.

### **3. Evaluación**

- Proyecto (100%)
- Tareas

#### **Fechas de presentación del proyecto:**

- **Primer parcial:** viernes de la semana 4
- **Segundo parcial:** viernes de la semana 8
- **Presentación de Proyecto Final:** viernes de la semana 11

#### **Criterios y escalas para la asignación de la calificación definitiva.**

- Menos de 6, es **NA**
- De 6 a menos de 7.5, es **S**
- De 7.6 a menos de 8.5, es **B**
- De 8.6 a 10, es **MB**

#### **Bibliografía:**

1. Wes McKinney, “*Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*”, 2nd Edition, O’Really Media, 2017.
2. Joel Grus, “*Data Science from Scratch: First Principles with Python*”, 2nd Edition, O’Really Media, 2019.
3. Jake VanderPlas, “*Python Data Science Handbook: Essential Tools for Working with Data*”, 2nd Edition, O’Really Media, 2023.

4. Andreas C. Müller and Sarah Guido, “Introduction to machine learning with Python a guide for data scientists”, O'Really Media, 2018.
5. Mendenhall, W. *Introducción a la Probabilidad y Estadística*, 14ª ed, 2015.